

Analyse de screening RNAi par HTS-Net

Ghislain Bidaut (ghislain.bidaut@inserm.fr)

IR Plateforme Cibi (CRCM's Integrative Bioinformatics)

Web: <http://cibi.marseille.inserm.fr>

Forge: <http://forgecrcm.marseille.inserm.fr/projects/cibi>

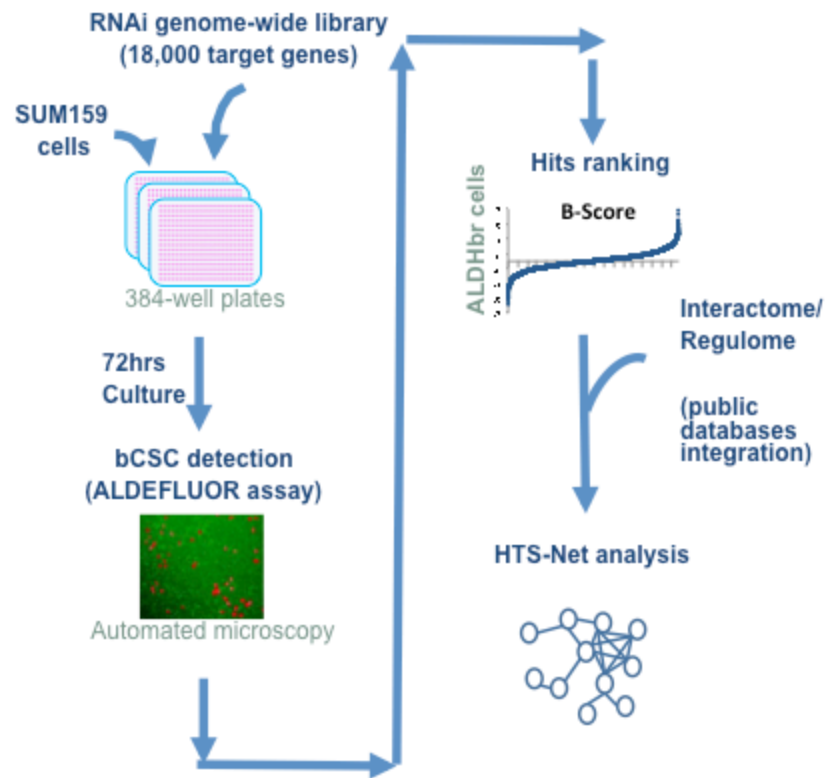


Screening RNAi

- Consistent à silencer individuellement l'ensemble des gènes d'un génome in vitro/in vivo par transfection de shRNAs/siRNAs.
- Permettent de comprendre fonctionnellement des phenotypes divers (pathologies, interactions hôte-virus, developpement, etc...)
- Les cellules étudiées sont marquées par GFP. L'abondance des cellules marquée est lue par microscopie associée à un traitement d'image automatique.
- Calcul d'un *z-score* associé à chaque gène, proportionnel à son impact sur l'abondance des cellules.
- Les gènes perturbés permettent de comprendre un phénotype de façon fonctionnelle.

Application au CRCM

Identification des gènes de la différenciation des cellules souches cancéreuses ALDH1 dans les tumeurs du sein. Projet porté avec Christophe Ginestier.



... Mais

... Comme pour beaucoup d'analyse haut-débit, on obtient de **longues** liste de gènes **sans causalité** ni **structure biologique** *apparente*.

- Nous proposons une **analyse réseau** comprenant
 - Des **interactions protéines-protéines** (PPI)
 - Des données de **régulation** (TF-ADN)
 - Les **scores du screening RNAi** étudié.
 - Une **visualisation** globale des sous-réseaux obtenus.

... pour replacer les résultats de screening RNAi dans un contexte biologique.

Données d'interactions

- Bases publiques d'interactions PPI

- Database of Interacting Proteins (*DIP*)
- Human Protein Reference Database (*HPRD*)
- Molecular Interaction Database (*Mint*)
- The Interologous Interactions Database (*I2D*)
- Human Proteinpedia

Total: 14423 gènes et 171146 interactions constituent *l'interactome*

Données de régulation

- **TF-TF Interactions**
 - TRANSFAC
 - The Atlas of Combinatorial transcriptional regulation (Ravasi *et al.*)
 - Database by Yu *et al.*
- **TF-DNA interactions**
 - TRANSFAC
 - ORegAnno
 - The Integrated Platform of mammalian Transcription Factors (ITFP)
 - PAZAR
 - The Regulatory Element Database (*TRED*)

Total: 9755 gènes et 68703 interactions constituent le *régulome*

HTS-Net

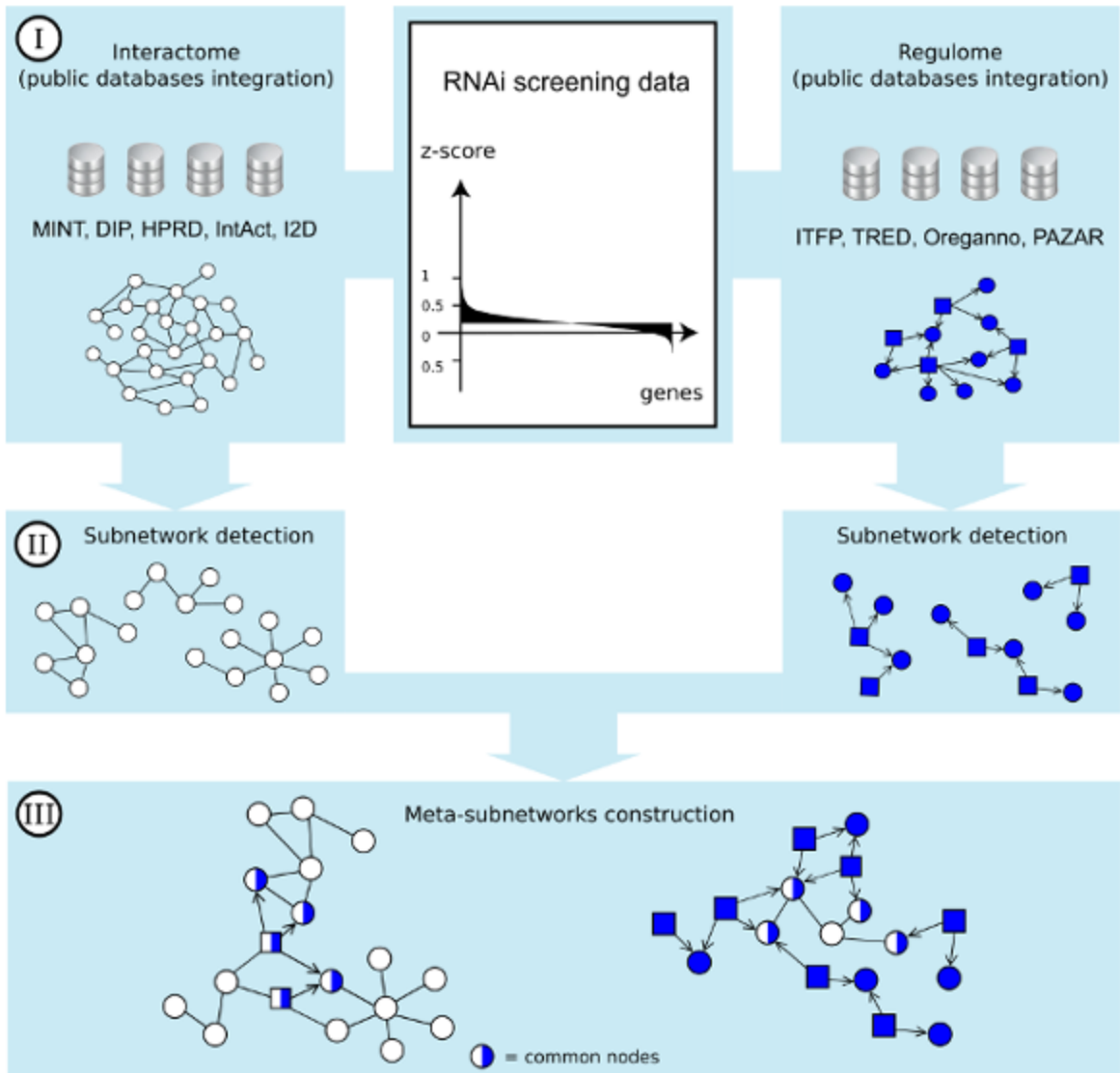
- Nous proposons un nouvel algorithme: **HTS-Net** (High-Throughput Screening – Network) pour **intégrer** les données **d'interactions** et les **données de screening**.
- Nous montrons les résultats obtenus sur des données publiques (**3 jeux** de données).
- Un pipeline HTS-Net a été installé et mis à disposition sur un portail *Mobyle* au CRCM.
- Nous avons comparé HTS-Net avec les programmes d'analyse réseau *CARD* et *HotNet2*.

Méthodes

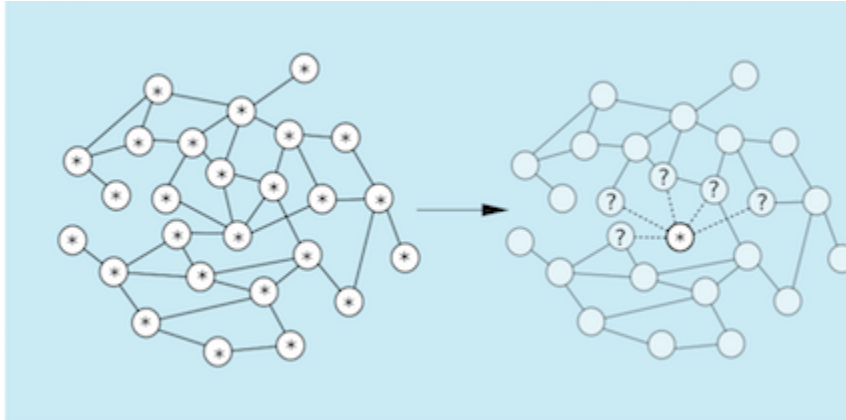
- Pour l'**Interactome** (Interactions PPI)
- Pour le **Régulome** (Interactions TF-ADN)
 - Superposition des z-scores sur un réseau d'interactions.
 - Extraction de sous-réseaux à z-score moyen élevé
 - Suppression des sous réseaux overlappant et validation statistique
 - Mélange aléatoire de l'interactome/regulome
 - Mélange aléatoire des z-scores

On valide donc statistiquement le **lien structurel z-score - réseau**.

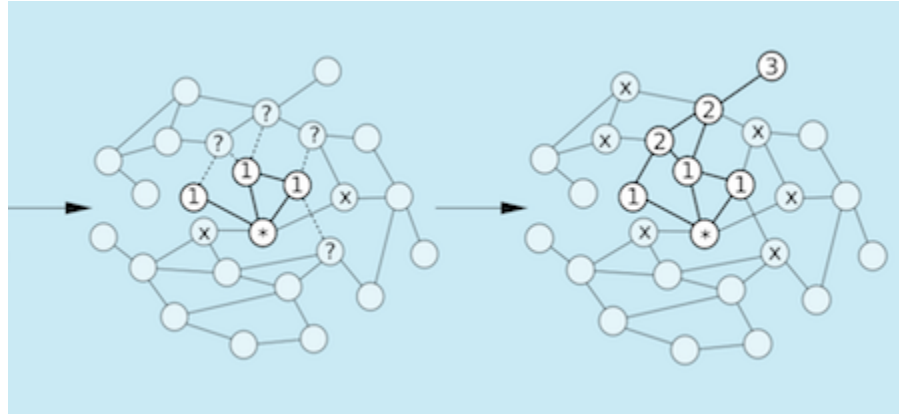
Pipeline HTS-Net



Identification des sous-réseaux (1)



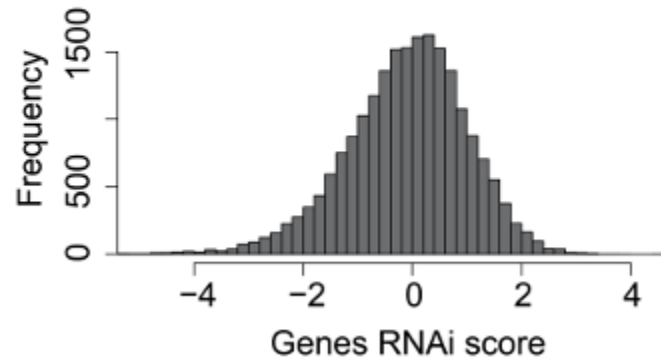
Identification des sous-réseaux (2)



Nature de l'analyse

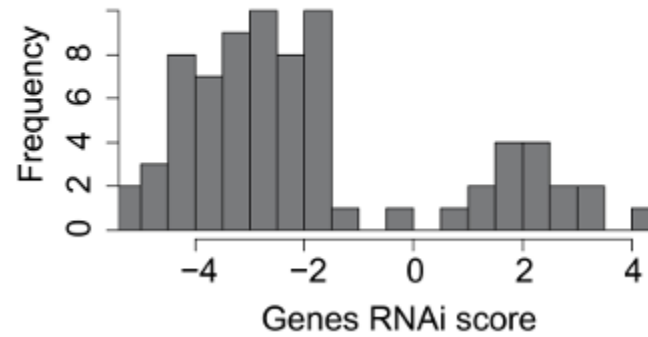
(A)

Tai Luciferase Gene Score - Whole screening



(B)

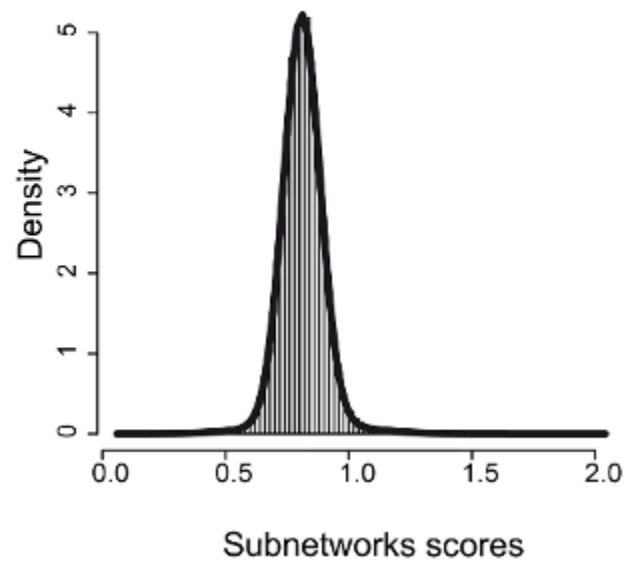
Tai Luciferase Gene Scores - HTS-Net



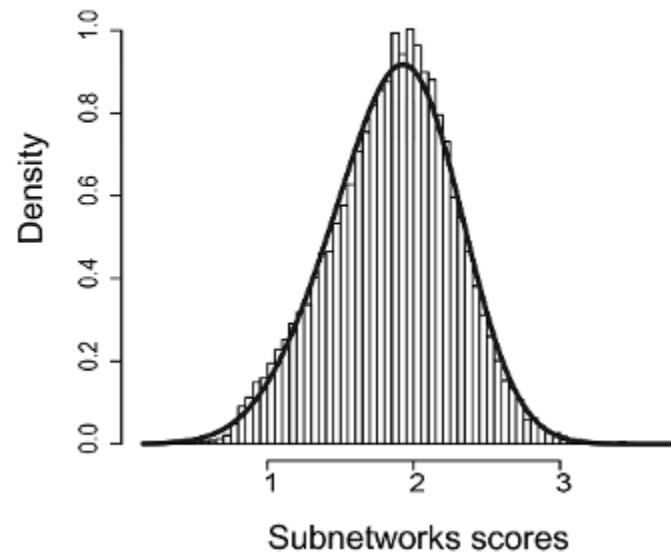
Validation statistique des sous-réseaux

1. Mélange aléatoire des interactions
2. Mélange aléatoire des scores

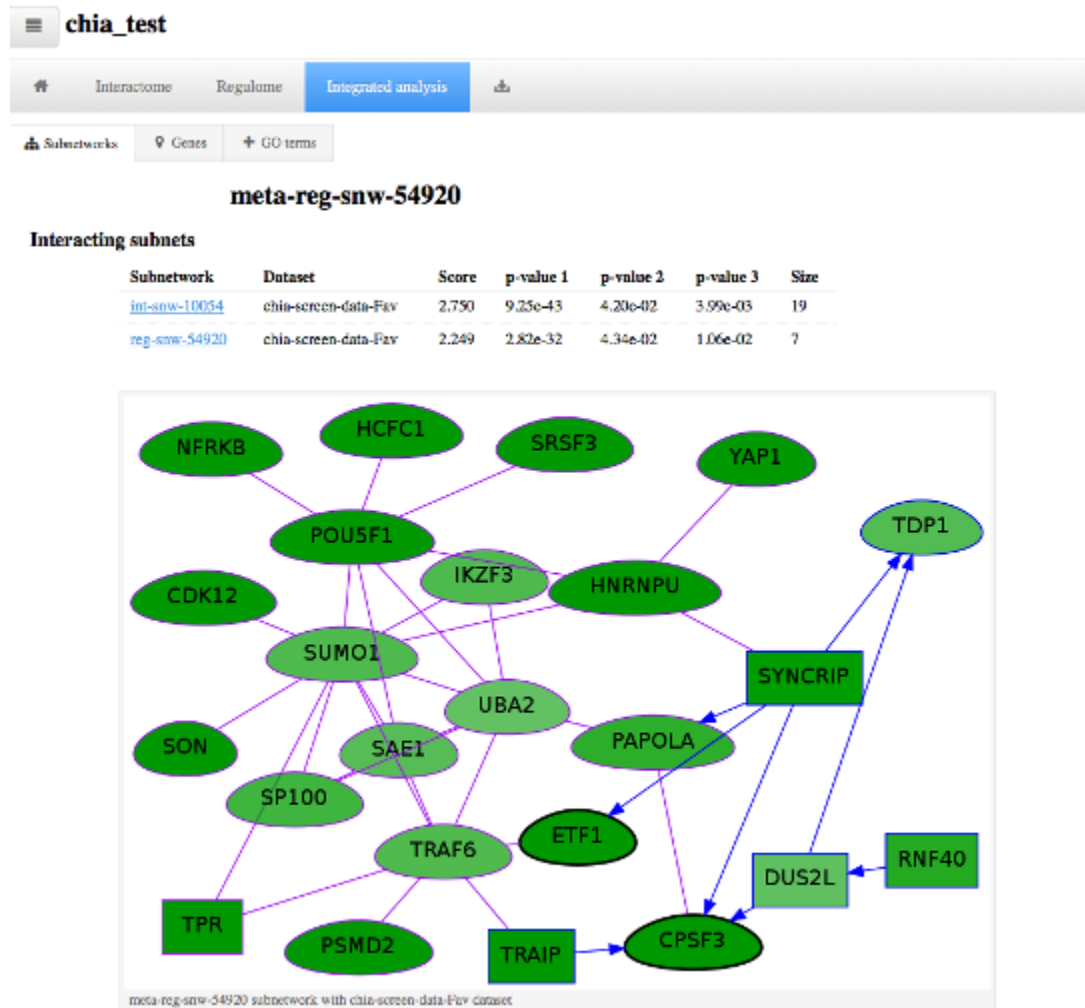
Type I random score distribution



Type II random score distribution



Exemple d'un sous-réseau obtenu



Pipelines existants

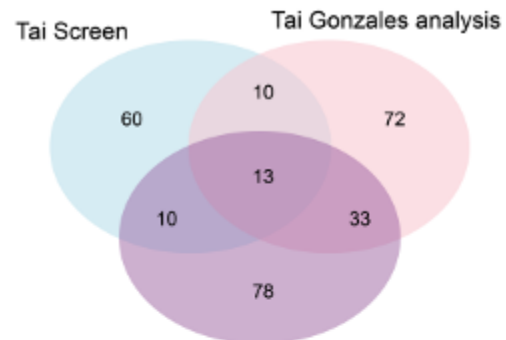
- *Méthode Gonzales & Zimmer*
 - Spécifique RNAi
 - Co-clustering des gènes sur la base de leur z-score et de leur proximité dans l'interactome
- *CARD*
 - Framework complet dédié au RNAi
 - Analyses bas niveau intégrée (traitement d'image et normalisation)
 - Analyse réseau: Superposition des hits sur le réseau et détection des "zones denses"
- *Hotnet2*
 - Spécifique pour la superposition de données de mutations ADN (SNPs, variants) sur l'interactome et détection de sous-réseaux contenant des *mutations* de façon significative

Comparaisons sur le jeu de données Tai

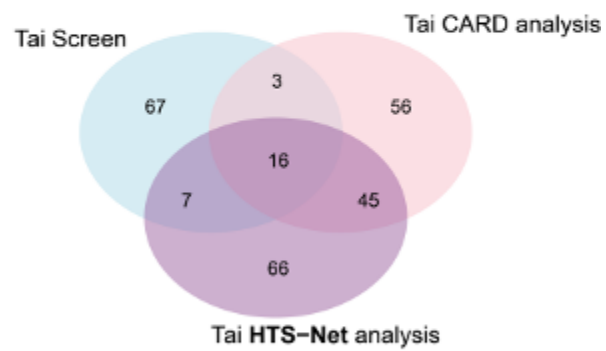
- *Gonzales & Zimmer*
 - Pas de code dispo, on teste sur les résultats de la publication.
 - 128 gènes identifiés. EIF3M a été identifié mais pas PIP5K1A ni HACD3
- *CARD*
 - Quelques résultats communs avec HTS-Net (COPA, COPB1)
 - Pas d'identification des mécanismes de régulation
- *Hotnet2*
 - Pas d'identification de gènes communs
 - Installation complexe
 - 40 heures de calcul seulement sur HPRD

Comparaison autres méthodes

A



B



C

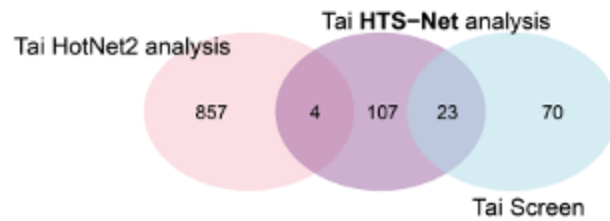


Tableau récap

Method	Type	Interactome	Statistical validation	Cost value/subnetwork score	Specificities	Disk footprint	Running time with default parameters.
HTS-Net	Web server+Source code executable	HPRD, BIND, DIP, I2D,IntAct, MINT, ProteinPedia, TRED, ITFP, TRANSFAC (internal use only), OregAnno, PAZAAR	Topology of gene network	Maximisation of subnetwork score	Triple statistical validation, regulation analysis on regulome	500Mo	1h48
Gonzales & Zimmer	Not provided	STRING	Co clustering on network proximity and score	Maximisation of subnetwork score	Approach based on clustering	Unknown	Unknown
CARD	Web Server	HPRD, BIND, BioGrid	Number and size of subnetworks	Gene connectivity	Complete web server that includes low-level RNAi data processing	Unknown	40minutes
HotNet2	Source code to be installed	HPRD, iRefIndex	Number and size of subnetworks	Diffusion and maximization of subnetwork score	Initially designed for mutation analysis.	60Go	40+ hours

Avantages d'HTS-Net

- *Pourquoi ne pas utiliser GSEA (Gene Set Enrichment analysis) ?*
 - Analyse liée aux 'sets' de gènes. C'est une mesure d'enrichissement.
 - Pas de découverte de nouveaux pathways
- *Pourquoi ne pas utiliser Ingenuity ?*
 - C'est une 'Boite noire' (liste des interactions non disponibles)
 - Conçu et limité aux **analyses transcriptomes**
 - Licence **propriétaire**
 - Payant pour **utiliser** le programme
 - Payant pour **retrouver ses données**

Analyse Chia *et al.* (Nature 2010).

- Déterminer les acteurs génétiques qui gouvernent l'identité des cellules souches.
- Screening primaire par pool de 4 siRNA pour chaque gène.
- Selection de gènes par mesure de la GFP -> z-score > 2.
- Screening secondaire apres déconvolution des siRNA.

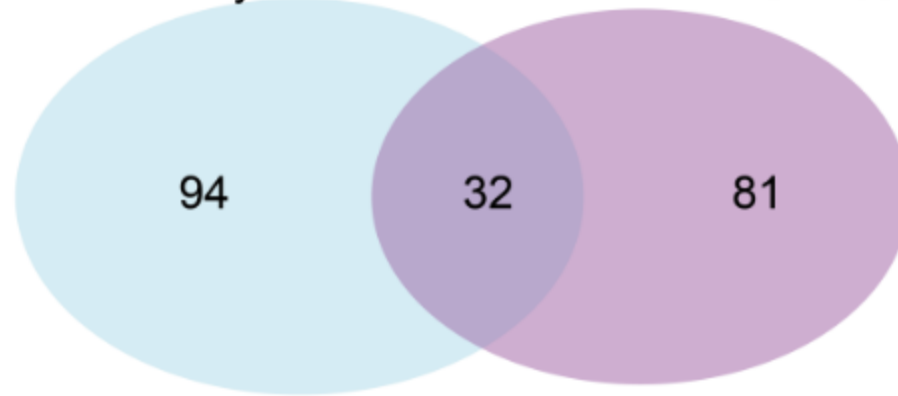
Chia a identifié au final **126 gènes** et nous en avons identifié **113 avec HTS-Net**.

- z-score élevé = GFP élevée = marqueur de non-différentiation élevé.

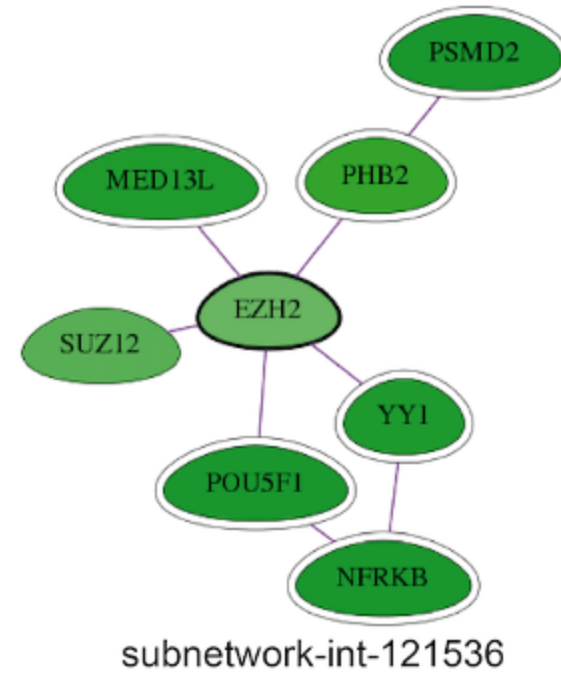
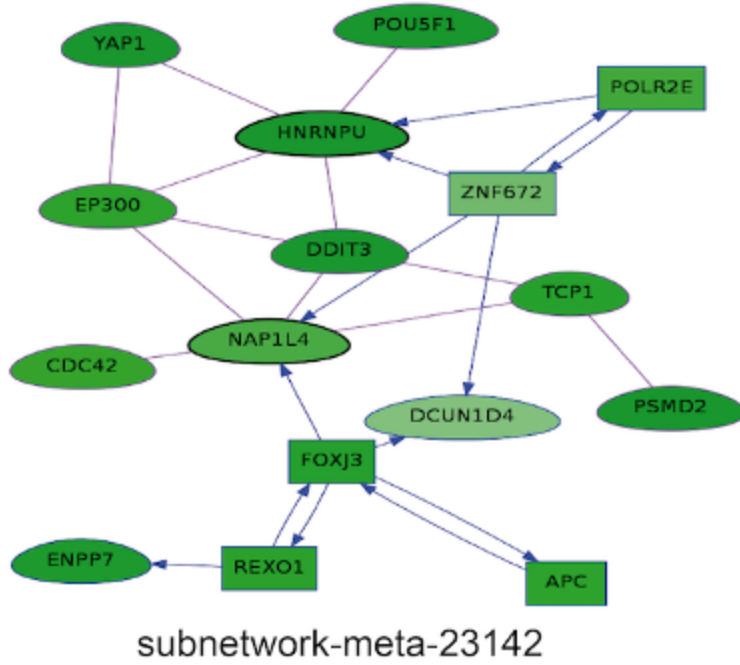
A

Chia Secondary Screen

Chia **HTS-Net** analysis



A: Analysis by Chia et al.



Résultat *HTS-Net* sur Chia

- Chia a identifié les régulateurs transcriptionnels PRDM14, NFRKB et YAP1.
- Ils ont été identifiés avec HTS-Net.
- PRDM14 = 7 sous-réseaux de régulation.
- NFRKB = 3 méta sous réseaux.
- YAPI1 = 12 méta sous-réseaux.
- Autre gènes, non détectés par Chia, liés à la régulation de la chromatine.
- PRC2
- EZH2 (identité cellulaire durant différenciation)
- SUZ12 (différenciation, cancer)

Analyse Wolf *et al.* (Breast Cancer Res)

Wolf et al a testé la propriété de cellules souches cancéreuses (CSC) à former des mammosphères avec un screening shRNAi. Ils ont comparé l'analyse avec les meme cellules sous conditions adhérentes.

Ensuite, il a sélectionné les gènes liées à la formation des mammosphères par seuillage de z-scores, et p-values sur les deux screenings. Obtention finale de **392 gènes**. Ensuite, l'analyse a porté sur les gènes ATG4A et ATG4B (autophagie).

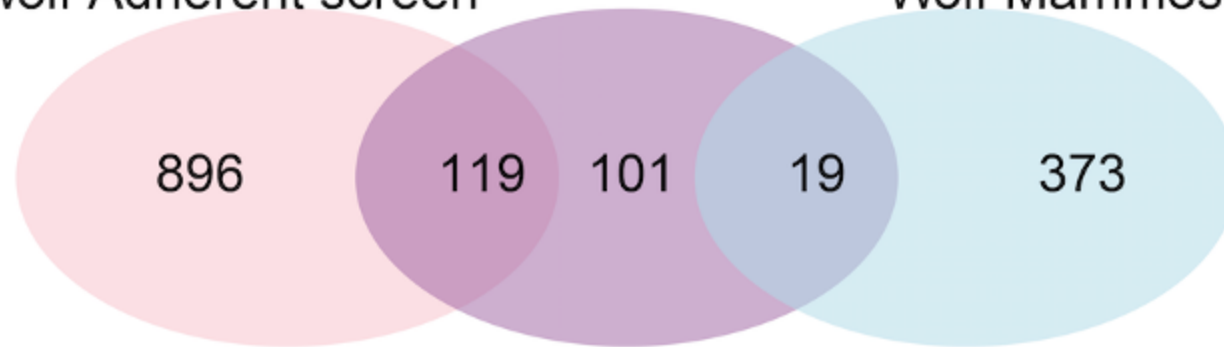
- z-score 1 = Mammosphère
- z-score 2 = Adherent

L'analyse HTS-Net a porté sur $\log_2(\text{Mammosphère}/\text{adherent})$ et a permis d'identifier les sous réseaux qui différencient les cellules adhérentes des cellules qui développent des mammosphères.

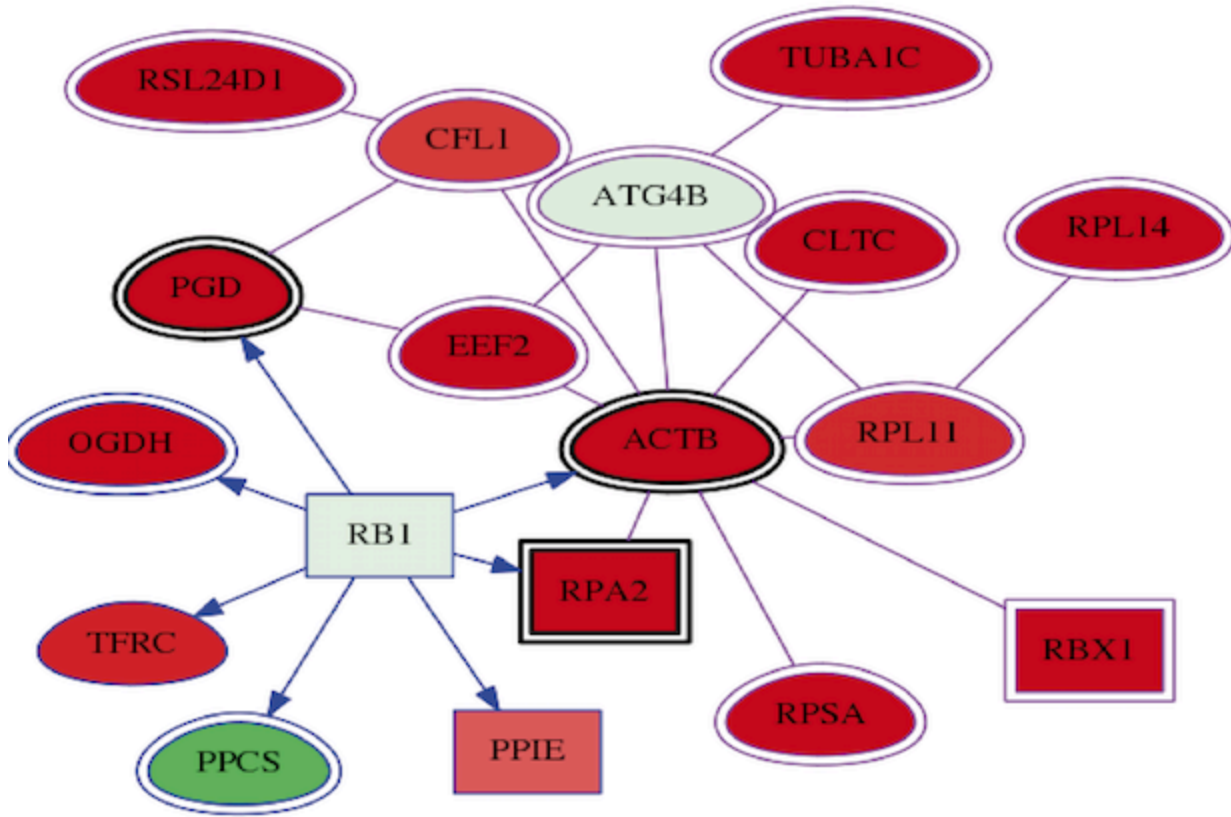
Wolf HTS-Net analysis

Wolf Adherent screen

Wolf Mammosphere Screen



B: Analysis by Wolf et al.



subnetwork-meta-23192

Résultats *HTS-Net* sur Wolf

239 gènes ont été détectés. ATG4B est détecté dans un sous réseau unique.

L'analyse en enrichissement GO montre que plusieurs fonctions liées au cycle cellulaire et différenciation cellulaires ont été détectées spécifiquement par HTS-Net.

- 'mitotic cell cycle': 32/415 (HTS-Net) 15/415 (Wolf)
- 'DNA Damage response': (13/68 (HTS-Net) 1/68 (Wolf)

Analyse Tai *et al.* (Cell 2009)

Tai a mis en place un screening RNAi pour la découverte des protéines qui permettent la réplication du virus HCV dans l'humain.

Le screening est fait sur la lignée cellulaire Huh7/Rep-Feo en répliquant sur une bibliothèque de 21,094 siRNA.

Le z-score est lié au fold change sur la luciférase (marqueur de réplication HCV).

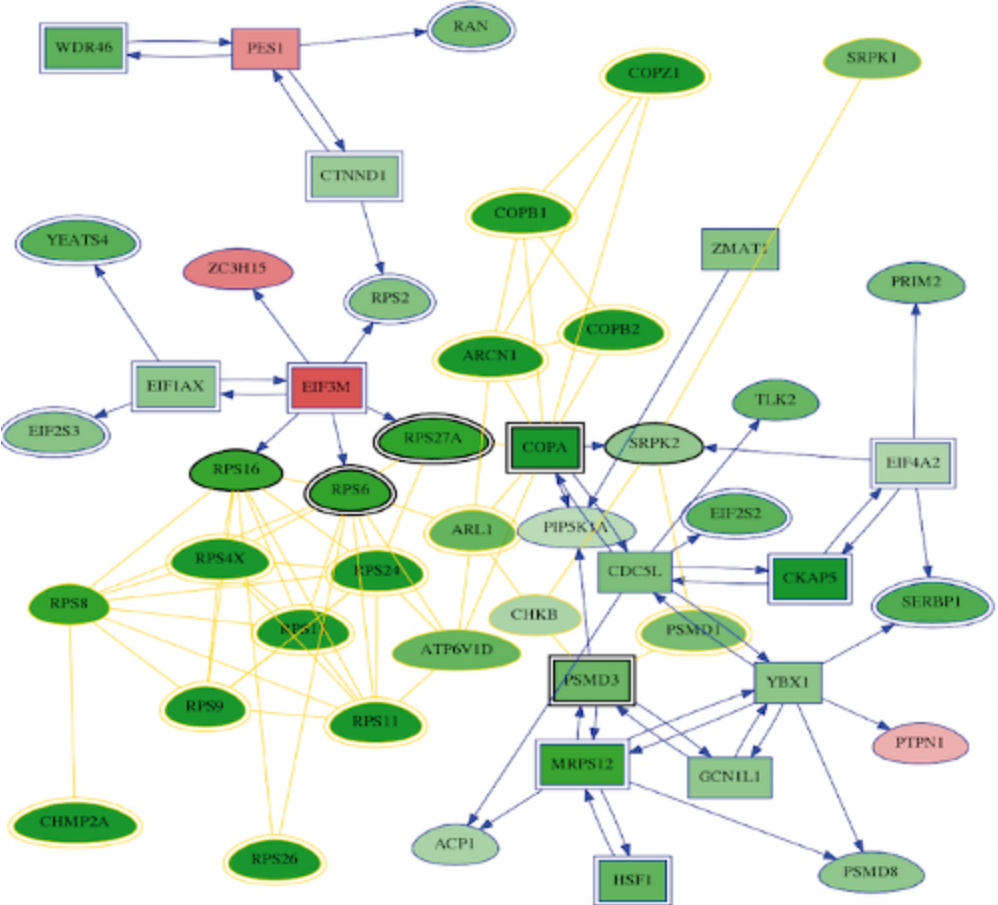
Identification de 93 gènes par Tai.

L'analyse *HTS-Net* a permis l'identification de 141 gènes, dont 23 communs avec Tai.

C



C: Analysis by Tai et al.



subnetwork-meta--1120

Résultats *HTS-Net* sur Tai

Les études Tai et HTS-Net ont permis l'identification de gènes du complexe COPI.

- HTS-Net a permis la confirmation de COPA, COPB1, COPB2, COPG1 et COPZ1. Ce complexe a été confirmé par Tai comme étant essentiel à la réplication HCV. CDC42 a également été confirmé par HTS-Net.
- BRCA2 a été identifié spécialement par HTS-Net, ainsi que plusieurs gènes liés au ribosome.
- Le réseau snx-7133 est régulé par EIF3M, requis pour la synthèse de protéines
- PIP5K1A a un score moyennement significatif (-1.8) mais a été retenu par *HTS-Net*.
- HACD3 est un régulateur de la réplication virale. Il a été spécifiquement détecté par *HTS-Net* malgré son score faible (0.875).

GO Enrichment - 'viral transcription' - 'viral life cycle'

Implémentation sur MobyLe

- Mise à disposition sur un serveur web <http://htsnet.marseille.inserm.fr>

Conclusion

- Mise en place d'un pipeline d'analyse réseau, *HTS-Net*, dédié au screening RNAi
- 3 jeux de données analysés par HTS-Net. Identification des gènes d'intérêt plus de nouveaux candidats.

Comparaison aux programmes existants:

- **Meilleures performances** que CARD et HotNet
- Extraction des mécanismes de **régulation**
- **Large base d'interactions**

Publié: Rioualen C, Da Costa Q, Chetrit B, Charafe-Jauffret E, Ginestier C, Bidaut G. HTS-Net: [An integrated regulome-interactome approach for establishing network regulation models in high-throughput screenings](https://doi.org/10.1371/journal.pone.0185400). *PLoS One*. 2017 Sep 26;12(9):e0185400. doi: 10.1371/journal.pone.0185400

Remerciements

Financement par projet INCa PL-Bio:

- Quentin Da Costa, MS (Maintenant IPC)
- Claire Rioualen (Maintenant au TAGC)
- Bernard Chetrit (plateforme DISC)
- Christophe Ginestier, PhD
- Emmanuelle Charafe, MD, PhD.

Plateforme Cibi

- Maxima Garcia (développeur initial ITI, maintenant SKU)
- Samuel Granjeaud, PhD
- Lucie Khamvongsa, MS (En Thèse avec J. Van Helden)
- Benoît Goutorbe (Stage M2 avec Alphabio)
- Arnaud Guille (Stage M2, maintenant IPC)

Merci de votre attention !

Questions ? Remarques ?

- Maintenant (ou plus tard) à ghislain.bidaut@inserm.fr
- Demandes d'analyse: Voir le site Cibi: <http://cibi.marseille.inserm.fr>.



- CRCM's Integrative BioInformatics